Pergamon

0042-6989(95)00336-3

Vision Res., Vol. 36, No. 16, pp. 2515–2529, 1996

Printed in Great Britain
0042-6989/96 $15.00 + 0.00

Preattentive Perception of Elementary Three-dimensional Shapes

JENNIFER Y. SUN,*† PIETRO PERONA*†

Received 2 May 1995; in revised form 23 October 1995

Experiments in which a single target pattern is discriminated from multiple background distractors show that certain shaded, two-dimensional (2-D) stimuli consistent with a top-lit, polyhedral interpretation can be processed fast (<80 msec) and in parallel. Unshaded line drawings of the same shapes, however, are processed serially. Strong pop-out asymmetries and control experiments involving shaded patterns that do not have familiar 3-D interpretations suggest that such fast, parallel processing is dependent upon perception of 3-D shape. Furthermore, this process can be influenced by contextual scene information, in a manner that is dependent upon whether the additional cues contribute to the perception of a consistent 3-D scene. Copyright © 1996 Elsevier Science Ltd.

Shape-from-shading Parallel processing Asymmetry Context effects

INTRODUCTION

When we look at the world around us, we see it as three-dimensional (3-D). No matter if we are viewing a physical 3-D scene, or even just a black-and-white photograph, our sense of shape is compelling. It is clear that, devoid of stereo disparity and color, gray-level images nonetheless contain many cues from which we can build a 3-D percept—luminance edges, shading gradients, occlusion contours, cast shadows, to name a few. Yet which and how, and along what time course, are these cues combined in the process of 3-D perceptual build-up? In our study, we investigate the pre-attentive processing phase of 3-D perception.

The classical studies of preattentive vision have dealt mainly with visual features of the one or two-dimensional (2-D) world. Typical stimuli included line edges, color, motion, as well as textons and various 2-D shapes. These patterns were sometimes even presented with stereo disparity, but the stimuli were typically neither displayed nor perceived as 3-D shapes (Beck, 1966, 1967, 1982; Olson & Attneave, 1970; Julesz, 1975, 1984; Treisman & Gelade, 1980). More recently, researchers have been turning their attention to the realm of 3-D shapes, and have found, to many's surprise, that preattentive vision does not appear to be constrained to operate only with two-dimensional objects (Ramachandran, 1988; Enns & Rensink, 1990, 1991; Braun, 1990, 1993; He & Nakayama, 1992; Kleffner & Ramachandran, 1992; Sun & Perona, 1993). There is also recent clinical evidence that shape interpretation occurs very early on in the visual processing hierarchy (Symons *et al.*, 1993). In particular, response-time experiments by Enns & Rensink (1990, 1991), showed that polyhedral targets differing from their distractors in their perceptual 3-D shape "pop-out" with the characteristics of preattentive processing. Moreover, Braun (1990, 1993), using a double-task method, showed that smoothly shaded circular stimuli that resemble spherical bumps or indentations give preattentive pop-out based on 3-D perception.

These results give rise to the following questions:

1. What are the relevant features in these stimuli that allow them to be processed in parallel as 3-D shapes? More specifically, is it the shading itself that is important or is it actually the edge boundaries created by the shaded regions?
2. Is the crucial computation performed locally, e.g. on corner junctions, or is it performed globally upon the entire shape?
3. Is this process a "hard-wired", local and bottom-up process, or can it be influenced by global and/or contextual information?

The first goal of this paper is to measure 3-D shape pop-out with an experimental paradigm that involves controlled display times and masking. This method will allow us to both verify Enns & Rensink's results as obtained from response time experiments, and to compare our results with those from previous pop-out and texture segregation experiments involving a similar paradigm (Bergen & Julesz, 1983; Kröse, 1987; Gurnsey & Browse, 1987; Nothdurft, 1991).

*California Institute of Technology 116-81, Pasadena, CA 91125 U.S.A.
†Universita' di Padova, Italy.
‡To whom all correspondence should be addressed [*Email* jensun@-vision.caltech.edu].

The second goal of this paper is to investigate the three questions raised above using two separate sets of experiments. The following section will deal with the nature and spatial extent of the critical feature. We will be comparing performance on the shaded polyhedral pattern found to elicit pop-out by Enns & Rensink (1990) with that of other patterns differing in specific characteristics of shading, contour, or orientation. Section 3 will address the importance of contextual information through experiments that contrast performance on displays that contain 3-D contextual information with those that do not.

PROCESSING OF SHADED PATTERNS

Methods

Subjects. Five female subjects and five male subjects, all between the ages of 18 and 40 yr, participated in the experiments. Subjects had normal or corrected-to-normal vision by self-report. All subjects were naive, except for one female subject.

Apparatus and stimuli. Images were generated on a Silicon Graphics Indigo with an 8-bit graphics display and a 16 msec screen refresh rate. Monitor dimensions were 221×295 mm, with a resolution of 3.47 pixels/mm.

Stimulus screens were viewed binocularly at a distance of 100 cm. Each stimulus screen contained 3, 12 or 24 items of display, with each item spanning approximately 1.5 deg of visual angle. In screens with 12 and 24 items, spacing between items was approximately 3 deg, measured from the center of one item to the center of its nearest neighbor, with an additional random jitter of up to 0.3 deg. For screens of three items, the separation was larger, approximately 7.5 deg, so as to maintain a comparable maximum eccentricity for all display sizes. Gray levels were produced using equal RGB pixel values, ranging from 0 to 256. Stimulus screens had a background RGB value of 80, which gave a measured background luminance of 0.84 cd/m^2. Shaded stimuli consisted of three regions, each of a different gray level. The corresponding RGB pixel values were 40, 180 and 256. Line stimuli were drawn with RGB values of 0.

Procedure. We used a two-alternative forced-choice (2AFC) stimulus onset asynchrony (SOA) paradigm with masking. Stimulus display times ranged from 16 to 400 msec depending on the task, and were followed by a blank inter-stimulus interval (ISI) time of 0, 16 or 26 msec and a 200 msec mask, an example of which is shown in Fig. 1(B). After the mask has disappeared, subjects were asked to report the presence or absence of the target pattern. Within each experiment, the target pattern was the 180 deg rotation of the distractor pattern. One target was present at random among multiple distractors in 50% of the trials. Target-present trials and target-absent trials had the same total number of patterns. Target position was also randomized, but was constrained to positions of 6.5 deg of eccentricity or less. Each experimental session consisted of about 1000 trials, presented in blocks of 35 trials, with number of items and duration of display held constant within a block. At the end of each experiment, subjects were asked to describe their perception of the stimuli. Subjects were trained until performance had stabilized, which typically took two training sessions.

Data analysis. Performance for each SOA duration was calculated using d-prime measurements (McNichol, 1972) derived from target-present and target-absent data, resulting in one psychometric curve for each subject in each condition. We make the assumption that, after an initial delay, the variance of the noise affecting the 2AFC decision is inversely proportional to the square root of the duration of the stimulus. Therefore, the psychometric curve was fitted, using a maximum likelihood fitting procedure, to the following model:

$$F(t) = \mathrm{Erf}\sqrt{\frac{t-m}{s}}$$

where m denotes the initial delay, and s is inversely proportional to the steepness of the function. The procedure involved obtaining a two-dimensional likelihood distribution over a range of values for parameters m and s. The pair of values giving the highest likelihood was used to generate a fit to the psychometric curve, and the SOA duration necessary to reach 75% accuracy was calculated from this fitted function. The spread of the likelihood distribution was used as a measurement of the goodness of fit. The mean SOA across subjects was calculated by averaging the fitted SOA of each subject, weighted by the goodness-of-fit estimate. For cases in which target detection was so difficult that performance did not saturate for even the longest display durations, the fitting was poor. In these instances, we estimated instead, to a 99% confidence level, the minimum SOA duration needed for 75% accuracy. This was done by normalizing the likelihood distribution and finding the a value up to which the area under the curve equalled 0.01. Minumum SOA durations across subjects were averaged and plotted with an asterisk (*) (see Fig. 7).

Experiments

Experiment 1(A) (Shaded cubes vs line patterns). In this experiment we attempt to establish whether the crucial component for pop-out and, arguably, pre-attentive shape perception is the oriented edges, the shading, or a combination of the two. We used for this investigation one shaded and two line stimuli previously used by Enns & Rensink in their response time experiments (Enns & Rensink, 1990, 1991). The shaded pattern consists of a shaded, Y-junction embedded in a hexagon [see Fig. 1(A)]. The distractors have an upright Y-junction and are typically interpreted as cubes sitting on a surface with lighting from above. The target has an upside-down Y-junction, and can be seen alternatively as a cube with its bottom side exposed and lit from below, or as a concave corner lit from above. Figure 1(B) shows a typical mask used in the shaded cubes experiments. The line cube and line Y-junction stimuli are shown in Fig. 1(C, D). Similarly to that for the shaded cubes, the masks

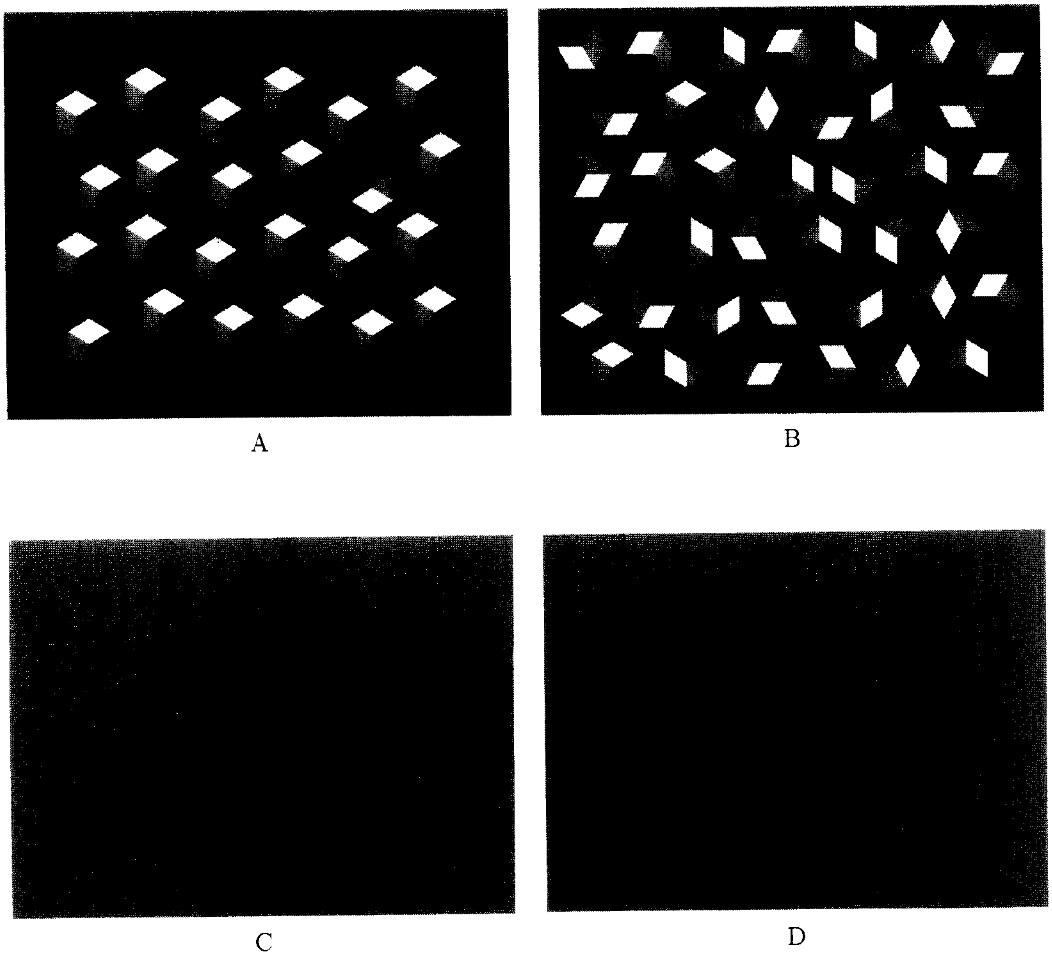


FIGURE 1. Top row shows a sample test screen (A) and a sample mask screen from the shaded cubes experiment (B). Bottom row shows sample test screens from the line cubes experiment (C) and the line Y-junctions experiment (D).

for these stimuli are composed of 0, 90, 180 and 270 deg rotated versions of the respective distractor pattern.

The average SOA durations necessary for 75% accuracy performance for the shaded cubes, the line cubes, and the line Y-junctions are plotted against the number of display items in Fig. 2. The average is taken across 7, 5 and 4 subjects, respectively. For the shaded cubes, performance is consistently fast across display sizes. The necessary SOA for processing is virtually independent of the number of distractors. When the data are fitted to a least-squares line, we obtain a slope of 0.8 msec/item, with a standard error of 0.6 msec/item. This result, with $P > 0.09$, is not significantly greater than a slope of zero. In contrast, necessary SOA durations for both the line cubes and the line Y-junctions increase dramatically with the number of distractors presented. The fitted slopes of 5.2 msec/item for the line cubes and 6.0 msec/item for the line Y-junctions are significantly greater than 0, with $P < 0.005$ for both cases.

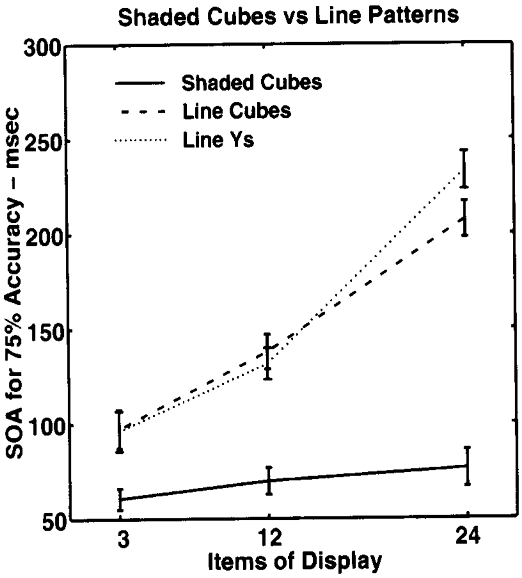


FIGURE 2. SOA for 75% accuracy for 3, 12 and 24 display sizes are shown for each of the three patterns: the shaded cubes, the line cubes, and the line Y-junctions.

This behavior, where the fitted slope differs significantly from zero, we will henceforth refer to as "serial", as opposed to the relative "parallel" behavior, as seen for the shaded cubes. This characterization of visual tasks as parallel or serial, based on the observed independence or dependence of performance on display size, has been made popular by Treisman and collaborators (Treisman & Gelade, 1980; Treisman, 1982; Treisman & Patterson, 1984; Treisman & Gormican, 1988). Not all researchers,

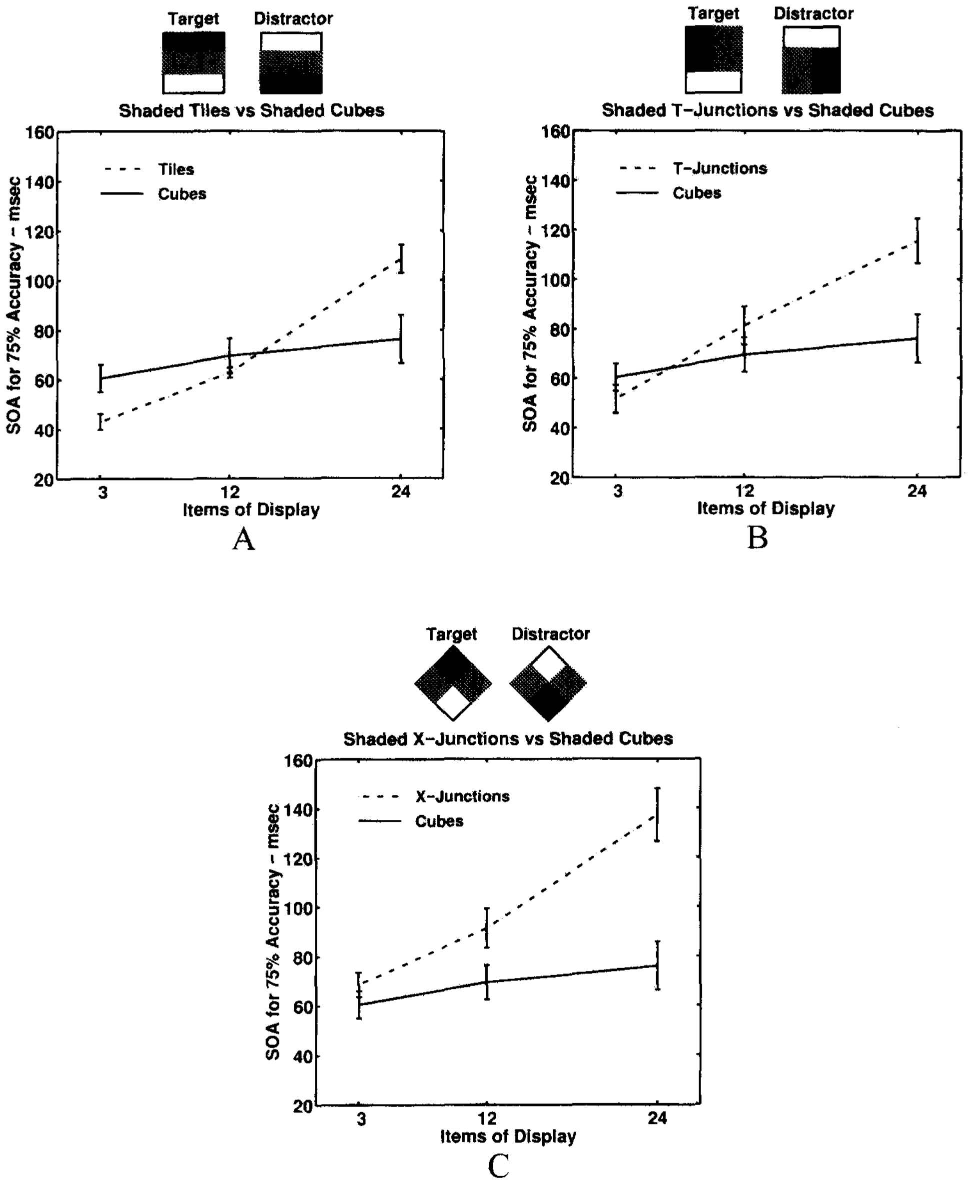


FIGURE 3. SOA for 75% accuracy for three shaded control patterns, shaded tiles (A), shaded X-junctions (B), and shaded T-junctions (C), are shown in comparison with that of the shaded cubes.

however, accept that there is a clean distinction; some prefer the notion that different degrees of task difficulty require different degrees of attention, resulting in a continuum of performance slopes (Duncan & Humphreys, 1989; Wolfe *et al.*, 1992).

Experiment 1(B) (Shaded 2-D patterns). The results from Experiment 1(A) show that shaded cubes, as opposed to line cubes, may be processed fast and in parallel. To investigate whether this fast and parallel processing is actually related to the "3-D-ness" of the shaded cube, we used three other patterns that, while shaded in the same black, gray, and white tones as the cubes, do not have typical 3-D interpretations. If these "flat" patterns can also be processed in parallel, then presumably the shaded cubes may also be processed in parallel using 2-D cues only, and the apparent "3-D-ness" of the shaded cubes may have nothing whatsoever to do with parallel processing.

Figure 3 shows the plots for three such stimuli which we call, respectively, the shaded tiles, the shaded T-junction, and the shaded X-junction, in comparison with the plot for the shaded cubes. The tile and the X-junction patterns were originally used by Enns & Rensink (1990). Data were collected from four different subjects for the tile experiment and five subjects each for the T-junction and the X-junction experiments. For three items of display, performances for the tiles and the T-junctions are, if anything, even better than that for shaded cubes. However, when display size is increased, necessary SOA durations for all three 2-D patterns are more strongly affected than that of the shaded cubes. The fitted slopes for the tile, T-junction, and X-junction patterns are all

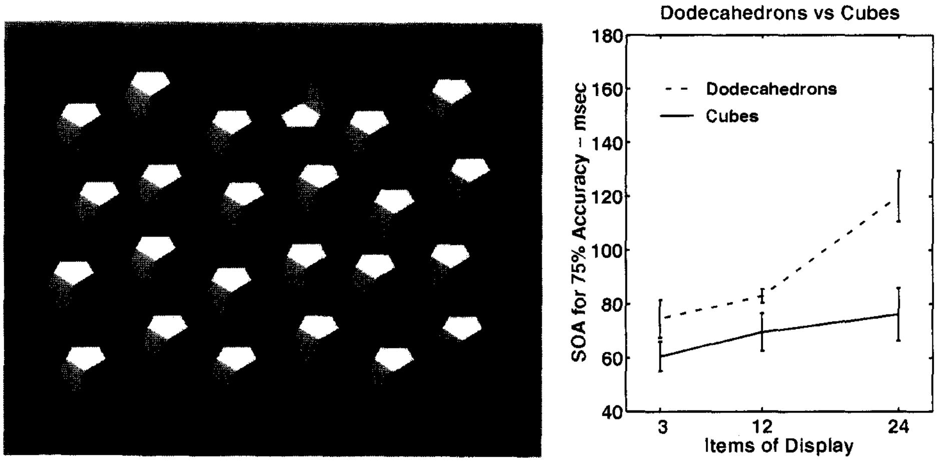


FIGURE 4. The graph on the right summarizes subjects' performance for this hexagonal shaded pattern (left). The performance is parallel for target-absent trials.

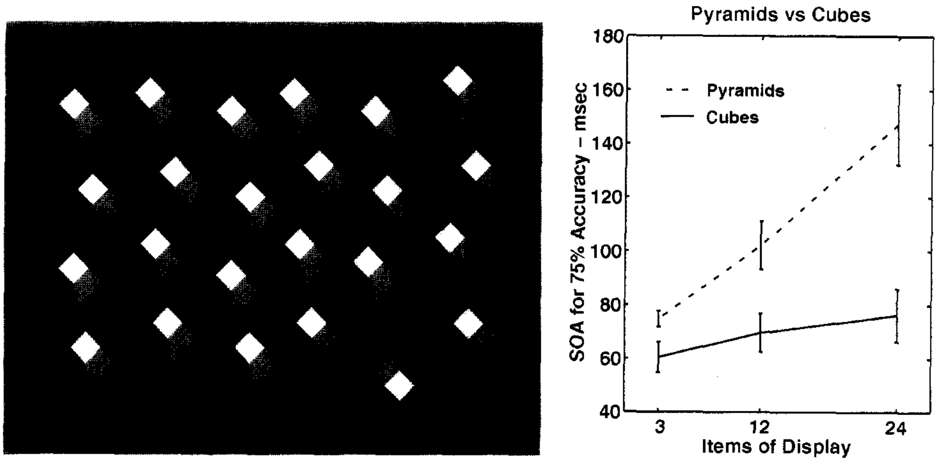


FIGURE 5. Stimulus display screen from the pyramids experiment is shown on the left, and the results are shown on the adjacent graph (right).

greater than 3.0 msec/item, and all significantly greater than 0, with $P < 0.005$. Enns & Rensink's response time experiments (1990) also show that the tile and the X-junction patterns are processed serially.

While the shaded cubes were recognized as convex, lit-from-above cubes without exception, none of the other patterns shown in this experiment prompted 3-D interpretations, except for the shaded tiles, which one subject voluntarily labeled as stairs.

Experiment 1(C) (Shaded Y-junctions). The normal orientation shaded cube pattern is composed of a central upright, shaded Y-junction and a hexagonal outline forming three arrow-junctions and three L-junctions. In Experiment 1(D), we ask whether this upright, shaded Y-junction is sufficient for fast, parallel 3-D perception, or whether the hexagonal outline that completes the figure of the cube is also necessary. To investigate the effect of the boundary contour, we embedded the Y-junction in three other outlines.

When the Y-junction is embedded in a 30 deg rotation of the outline of the cube, the resulting pattern can be interpreted as a dodecahedron (Fig. 4, left). Only two of the five subjects recognized it as such, however, while all described it as being less obviously three-dimensional than the shaded cube. Embedding the Y-junction in a diamond-shaped outline results in a pattern that resembles a truncated pyramid (Fig. 5, left). While all six subjects reported seeing the distractors as obviously three-dimensional at long display durations, they commented that the display was confusing and difficult to organize at short display durations. The results (Figs 4 and 5, right) show that necessary SOA durations for both these patterns are dependent on display size, with $P < 0.005$, reflecting the perception that these patterns look less consistently 3-D than the shaded cube.

We also embedded the shaded Y-junction in a circular outline (Fig. 6, left), resulting in a pattern that may be perceived as a 2-D pie chart. In itself, this pattern cannot be interpreted as a complete 3-D shape viewed under non-accidental conditions, but, if occlusion is postulated, it can be seen as one corner of a 3-D shape that is being viewed through a circular aperture. Data collected from

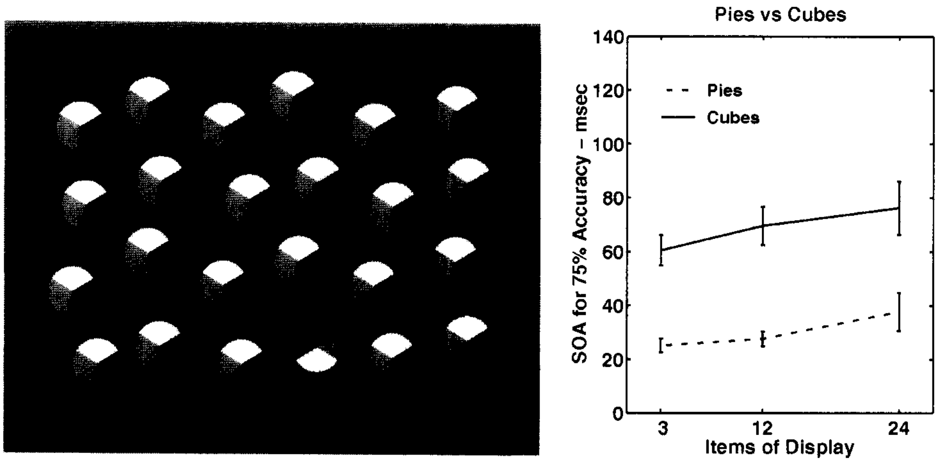


FIGURE 6. A stimulus display screen from the shaded Y-junction in circles experiment (left) and the corresponding performance (right).

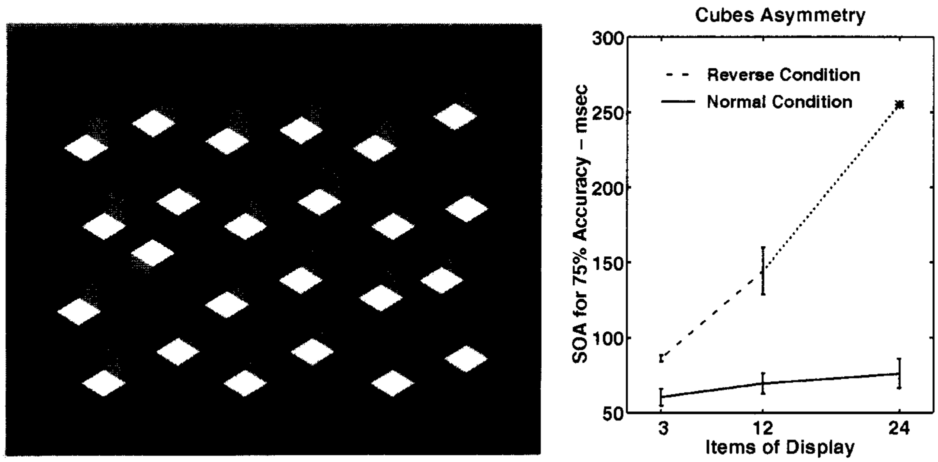


FIGURE 7. "Reverse" orientation shaded cubes are shown on the left. Results from this experiment are shown on the graph on the right.

five subjects revealed that this was the easiest task of all, even easier than the shaded cubes (see Fig. 6, right). The least-squares-fit line has a slope of 0.6 msec/item, and is not significantly greater than 0 ($P > 0.05$). While this pattern does not actually correspond to any complete 3-D object and may not look convincing 3-D when stared at on a static display, subjects reported that, during the experiment, the patterns looked like corners of cubes that are illuminated by circular spotlights or spikes sticking out through a curtain, vividly 3-D in any case.

Experiment 1(D) (Pop-out asymmetries). Nine different patterns have been investigated in the previous three experiments. Of these nine, only two, the shaded cubes and the shaded pies, appear to be processed fast and in parallel. It is often the case that when the target and distractor patterns of a parallel task are reversed, performance suddenly becomes serial. Such asymmetries are often used as a diagnostic in the search for primary visual features (Treisman & Gormican, 1988; Williams & Julesz, 1992). In this experiment we investigate whether our two parallel tasks will also become serial when the target and distractor patterns are reversed.

Figure 7 (left) illustrates the "reverse" condition of the shaded cubes experiment. Compared to the "normal" condition shown in Fig. 1(A), the target and distractor patterns are reversed. Since the target and distractor patterns are related by a 180 deg rotation, the "reverse" condition can be seen simply as a 180 deg rotation of the "normal" condition. We report a dramatic asymmetry in performance between the normal and reverse conditions. For the 24-item case of the reverse condition, performance did not saturate for any of the five subjects, even at the longest display duration. The estimated minimum SOA time (see Data analysis) averaged across subjects is plotted with an asterisk (*) instead. Since reliable 75% correct SOA times cannot be attained from the data, only the data points for the 3- and 12-item displays were used for line-fitting. The resulting fitted line has a slope of 6.5 msec/item, and is significantly greater than 0 ($P < 0.005$).

This asymmetry in performance is correlated with a perceptual asymmetry. While all subjects reported that

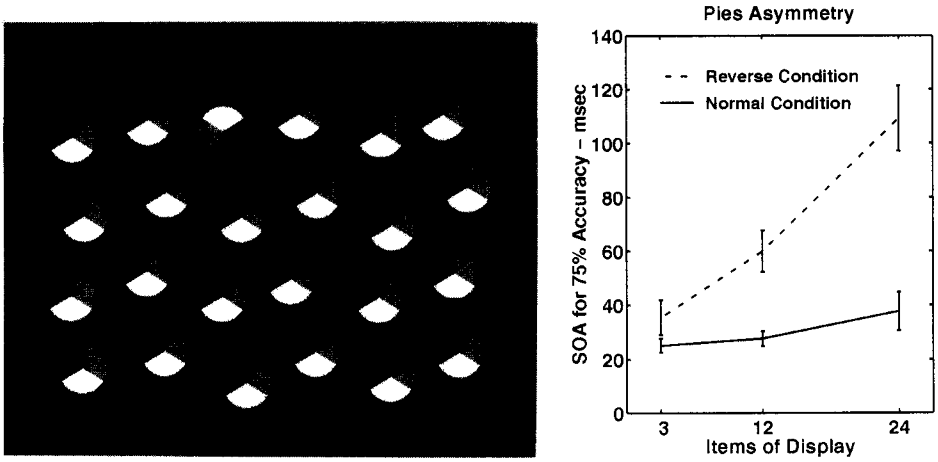


FIGURE 8. "Reverse" orientation shaded pies are shown on the left. Results from this experiment are shown on the graph on the right.

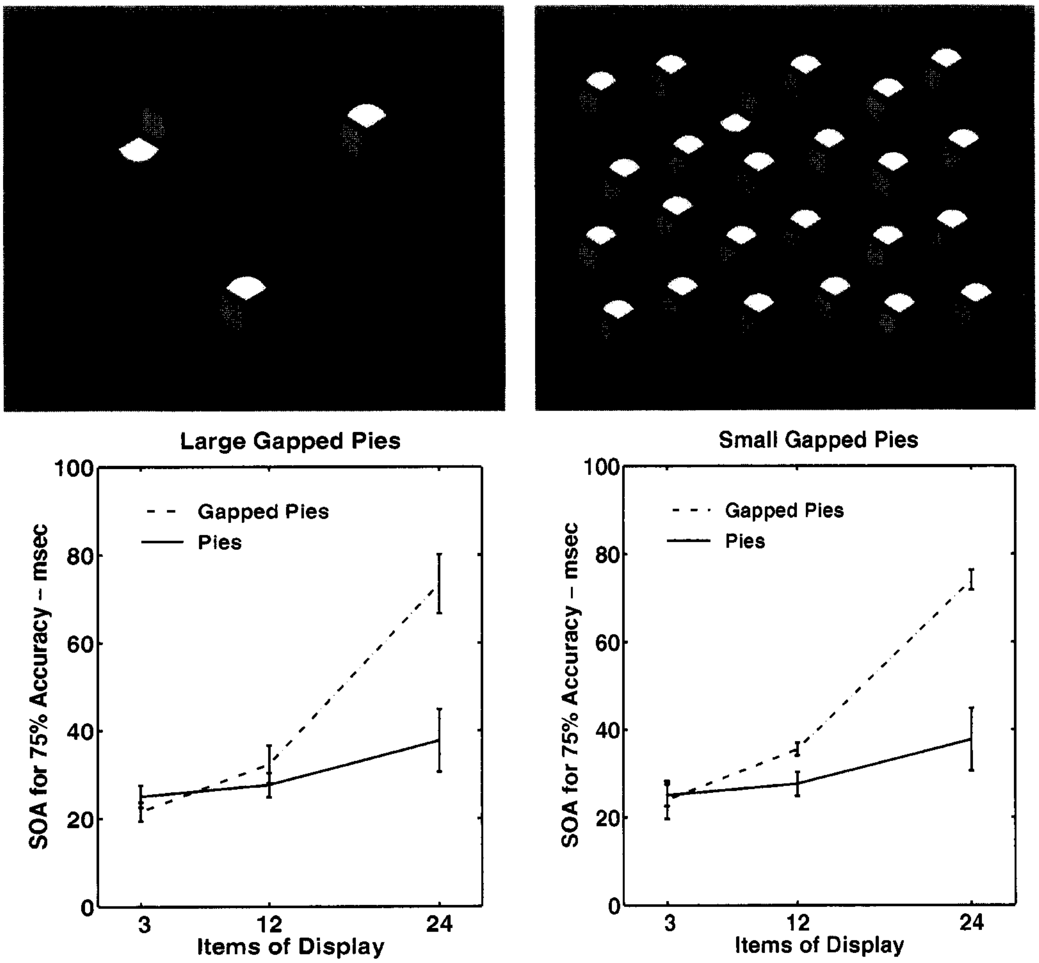


FIGURE 9. Experiments were conducted on gapped Y-junction in circles of two sizes (top). The corresponding performance graph is shown below each pattern.

the distractor in the normal condition experiment looked like lit-from-above convex cubes, and were strongly three-dimensional, none of the subjects reported 3-D interpretation for the background items in the reverse condition.

Figure 8 shows the "reverse" condition of the shaded pies (left) and the resulting performance, averaged across four subjects, in comparison with that of the "normal" condition shaded pies (right). The asymmetry is apparent. While performance for the normal condition shaded pies does not depend significantly upon display size, performance for the reverse-conditions shaded pies does, at a rate of 3.4 msec/item ($P < 0.001$).

Experiment 1(E) (Y-junction in circles). The results so far suggest that the normal orientation shaded Y-junction itself might be a salient feature in 3-D processing. To further explore this idea, we separated the shaded regions of the shaded pie from each other with a gap of about 0.2 deg. Two gapped patterns were used, one with shaded regions of the same area as the no-gap pies (Fig. 9, top left), and the other with a total area that was the same as the no-gap pies (Fig. 9, top right). Data were collected from four subjects for the large-gapped pies and two subjects for the small-gapped pies. Results show that the gaps make a significant difference (Fig. 9, bottom). Both the large and the small-gapped pies have slopes that are significantly greater than 0 ($P < 0.005$). Again, there is a correlation between 3-D interpretation and performance; subjects did not report 3-D interpretation for either sizes of gapped pies.

Discussion

Parallel processing of 3-D shapes. We confirm Enns & Rensink's finding that the shaded pattern consistent with the interpretation of a top-lit cube can be processed in parallel. We show that this parallel processing is also fast, requiring display durations of less than 80 msec, comparable to fast 2-D pop-out and texture segmentation processes (Bergen & Julesz, 1983; Kröse, 1987; Gurnsey & Browse, 1987; Nothdurft, 1991).

Our results support the idea that mechanisms computing some aspects of 3-D shape are involved in this fast, parallel processing. Both the results of experiment 1(B) (Fig. 3), which involved shaded patterns that do not have 3-D interpretations, and those of experiment 1(D), the asymmetry experiments (Figs 7 and 8), serve as evidence. Asymmetry in performance is seen for both the shaded cubes task and the shaded pies task. While the normal orientation tasks are distinctly easy for both, the reverse orientation tasks are significantly more difficult. Since the normal displays and reverse displays are merely 180 deg rotations of each other, they are entirely equivalent in 2-D terms. The vast perceptual difference between the two must therefore lie in their different 3-D interpretations, and a clear perceptual difference was in fact spontaneously reported by all our naive subjects. An analogous asymmetry was found also by Braun (1990, 1993) using smoothly shaded "bubble" stimuli.

Shading stimuli vs line stimuli. In their 1991 paper, Enns & Rensink show that, in a response time paradigm, tasks involving line cubes require search times that increase from approximately 500 to 700 msec as the display size is increased from one to six to 12 items. Other line patterns that do not have 3-D interpretations require search times that increase from 500 to more than 1000 msec. Their conclusion is that the visual system can process line arrow and Y-junctions preattentively, extracting 3-D structure rapidly and in parallel.

Our results, however, indicate otherwise (see Fig. 2). We find shading to be a crucial cue for driving this fast parallel process. When the shaded cube was replaced by an equivalent line drawing, performance was significantly compromised [experiment 1(A)] and became serial. We suspect that the difference between our findings is a consequence of our different experimental paradigms. Recent experiments (Sun & Perona, 1995, 1996) suggest that 3-D shape mechanisms driven by line drawings may be used for discrimination when display durations exceed 250 msec. In a response time experiment, where display durations are several hundreds of milliseconds, these mechanisms may be used to accomplish the task, thereby obscuring the the differences between the shaded and the line-drawing cases.

The shaded Y-junction. Results from experiments 1(D and E) indicate the normal orientation shaded Y-junction to be an important cue for preattentive 3-D processing. We find the shaded pies task to be even easier than the shaded cubes task (Fig. 6). This extreme ease of processing is disrupted, however, when the contingent shaded regions of the Y-junction are separated by a gap (Fig. 9).

One interpretation is that the simple and fast 3-D mechanism begins locally by processing the central Y-junctions. If no intrinsic surrounding corner junctions are present, it proceeds quickly to completion. This may apply to the case of the shaded pies. Subjects described this display as resembling convex corners seen through circular apertures. It is possible that the three rounded T-junctions on the surround are perceived as the results of occlusion, and are therefore not considered as intrinsic corners of the figure. In such a case, only the central Y-junction would need to be processed in order to achieve a 3-D percept.

When the surrounding corner junctions fit the configuration of a familiar 3-D shape, as in the case of the cubes, they are integrated with relative ease, at a small cost in processing time (Fig. 6). When the surrounding corner junctions cannot be easily integrated with the central junction to form a familiar shape, however, this basic mechanism fails. While a 3-D interpretation is possible for both the dodecahedron and the truncated pyramid stimuli, these patterns, containing T-junctions and L-junctions that cannot be readily perceived as resulting from occlusion, are accidental views of their possible physical interpretations (see Figs 4 and 5). The cube pattern, on the other hand, with a combination of arrow junctions and L-junctions on the surround, is a generic view of the prototypical cube (see Nakayama & Shimojo, 1992). Neither of these two other accidental views are as common and familiar as the generic cube view. This correlates with the subjects' perception that these shapes are unconvincingly 3-D or difficult to interpret during short durations of display.

A convex lit-from-above detector. Asymmetric pop-out is a topic of interest in the study of preattentive vision because it is indicative of the presence of a detector that is specialized for one of the two stimuli, but not both (Treisman & Gelade, 1980; Williams & Julesz, 1992). The asymmetry we find in experiment 1(D) suggests that this early vision 3-D mechanism also has a preferred stimulus, either a convex lit-from-above shape or a concave lit-from-above shape.

The feature detection theory proposed by Treisman and collaborators argues that if a single detector is used in a pop-out task, the task will be easier when the detector is specialized for the target, rather than the distractors. If the distractors are favored by the detector and the target is not, then the task is predicted to be more difficult. This line of reasoning would explain the asymmetry we observe by postulating the existence of a concave lit-from-above detector. However, since our subjects consistently reported easy perception of convex shapes when the background stimuli were convex lit-from-above patterns, and no perception of shape, be it concave or convex, when the background stimuli were concave lit-from-above patterns, we prefer the hypothesis that the convex lit-from-above shape computation is the one that is primarily subserved by this early vision 3-D mechanism.

Our preference might be accommodated by an alternative theory: Rubenstein & Sagi (1990) have suggested that asymmetries in pop-out performance have to do as much with the level of "noise" generated by the background as with the "signal" associated with the target. If shading patterns that promote a top-lit, convex percept (such as the normal orientation shaded Y-junction) are preferred by this 3-D mechanism, then, as distractors, they would generate a minimum of background noise. Among such a "quiet" background, the target could be spotted fast and in parallel. On the other hand, if the distractors are shaded patterns that do not promote the preferred interpretation (e.g., the upside-down shaded Y-junction), the background noise level would be high. To detect the signal generated by the target among such a noisy background would then require a serial search. Recent experiments (Sun & Perona, 1995, 1996) suggest that the most important cue for 3-D pop-out is reflectance, rather than 3-D shape or luminance. A percept of top-lit, convex shape leads to discounting of apparent luminance, resulting in a more uniform apparent reflectance. Top-lit, convex distractors would result in a quiet background and easy target detection, while distractors that do not promote such an interpretation would result in a noisy background and difficult target detection.

INFLUENCE OF CONTEXTUAL INFORMATION

The results from the previous pop-out experiments suggest the existence of mechanisms that can compute 3-D interpretations of shaded patterns composing a scene fast and in parallel. We also found that subject performance is highly correlated with reports of easy 3-D scene interpretation during short display durations. Is this 3-D pop-out based solely on local mechanisms, or can this process be influenced by global and/or contextual information? The following set of experiments investigate this question.

Methods

Experimental set-up. The same 2AFC SOA with masking paradigm was used for this second set of experiments. In experiments 2(A and B), 6, 12, 18 or 24 items of display were used. In test experiments, items ranged in size from 0.9 to 2.6 deg and were arranged according to size (See Fig. 11, top row and Fig. 12, left). In addition, one condition involved a background that suggested the context of a room. The background was displayed throughout the duration of the experiment. In control experiments, all items had the same size (1.5 deg of visual angle), and no background context cues were used. In experiment 2(C), 3, 6 or 12 items were displayed at an eccentricity of 4.3 deg of visual angle, with random jitter of up to 0.3 deg. In test experiments, the items were displayed within a wall frame that has a 3-D interpretation. In control experiments, the frame did not have a 3-D interpretation (see Fig. 15). Both 2-D and 3-D frames were displayed statically throughout the experiment.

Data analysis. Psychometric curves were fitted using the same method as in experimental series 1. SOA durations necessary for obtaining 75% accuracy were estimated for both test and control conditions for each subject. Individual improvements were combined by weighted averaging to give the mean improvement and an associated standard error. Data are presented in terms of mean improvement of performance under test conditions with respect to control conditions. The probability that improvement is significantly greater than zero is given as a measure of confidence.

Experiments

Experiment 2(A). We see from Fig. 10 that cubes of different sizes make a much harder task than cubes of all the same size, a size that is about the average of the cubes in the different-sizes condition. Necessary SOA duration for the largest display size increases more than 100 msec. In experiment 2(A), we investigate the condition in which the cubes are of different sizes, but are arranged in an orderly size gradient.

In one display condition, the arrangement mimics the effect of cubes sitting on the ground, receding off into the distance (Fig. 11, top-left). Figure 11 (bottom-left) plots the results in terms of improvement over control condition of same-sized cubes, with degradation of performance represented as negative improvement. There is an impairment of 20–30 msec.

When the perspective cues were reinforced by a background room context (Fig. 11, top-right), however, we found an overall trend of improvement. The confidence levels of improvement for the four increasing display sizes are 93%, >99%, >99% and 96%, respectively. Subjects reported that the perspective sizing enhanced the 3-D percept and made the task easier. In contrast, for what we call the ceiling perspective with room context (Fig. 12), which is a rather unusual viewing condition that does not fit with the apparent orientation of the shaded cubes, results compiled from three subjects show significant negative effects for the larger display sizes.

Experiment 2(B). Experiment 2(B) deals with the effect of contextual information on reverse orientation

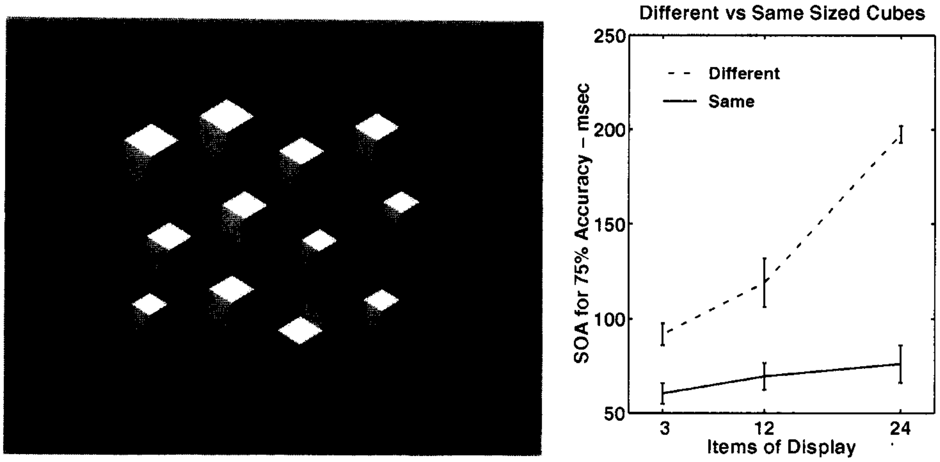


FIGURE 10. The performance of three subjects was tested using shaded cubes of different sizes (left). The results are shown in comparison to that of same-size shaded cubes on the right.

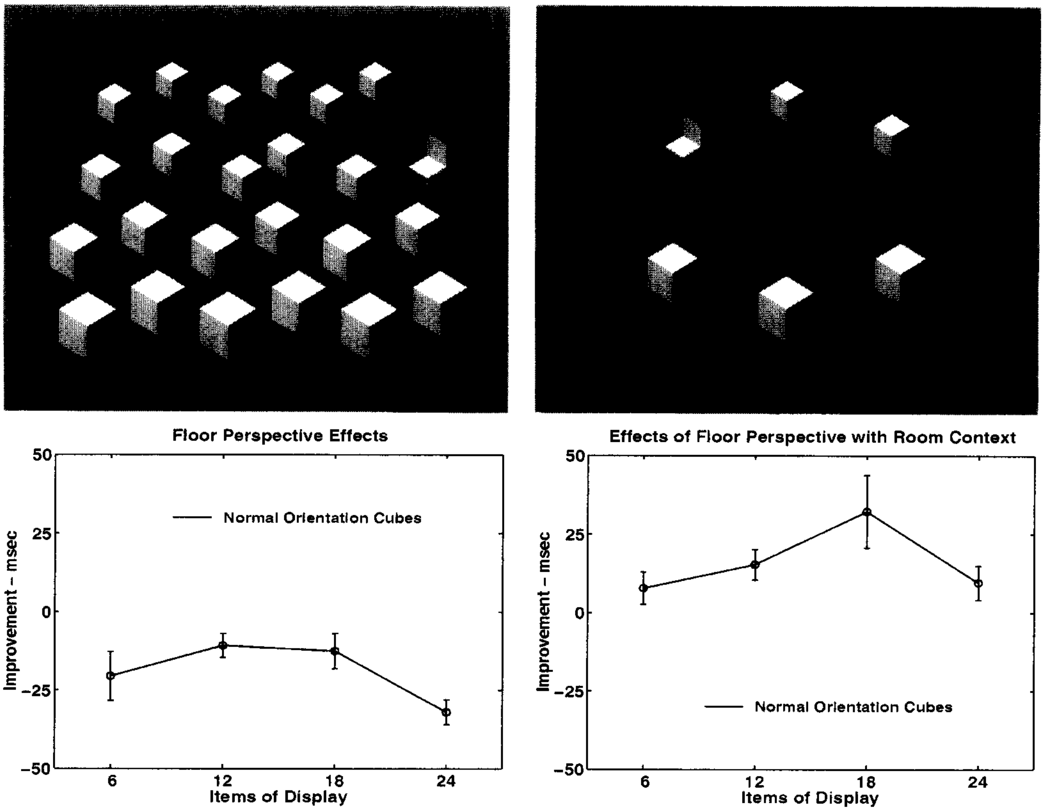


FIGURE 11. Top row shows normal orientation cubes arranged in floor perspective (left) and normal orientation cubes in a room context as well as in floor perspective (right). Bottom row shows the respective effects of these enhancements in terms of improvement. These plots reflect the data collected from four and three subjects, respectively.

cubes. Three subjects were tested on the ceiling perspective only experiment, and two subjects were tested on the one that included a room context.

When only perspective is used, we see a generally insignificant effect, except for the 6-item case (Fig. 13, left). When room context is added, however, the improvement becomes significant (confidence level at >99% for all display sizes), and especially large for the two larger display sizes (Fig. 13, right). Although we originally thought that these reverse orientation cubes would be best perceived as cubes hanging from the ceiling and, therefore, a floor perspective should hinder performance, some subjects reported that the floor perspective enhanced perception of the 3-D scene by allowing the stimuli that were previously difficult to interpret to be perceived as cubes balanced on a single vertex.

To test our hypothesis that the reverse orientation

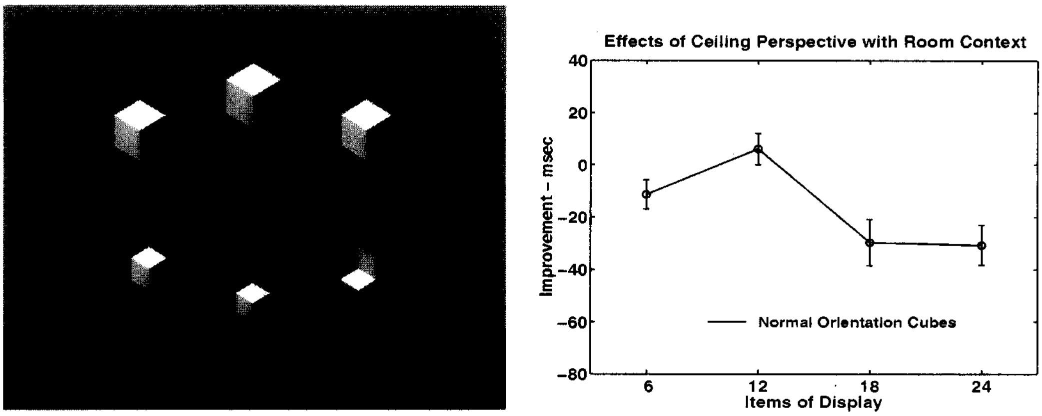


FIGURE 12. Normal orientation cubes were arranged in ceiling perspective and room context (left). The improvement of this condition over the standard task is shown on the right.

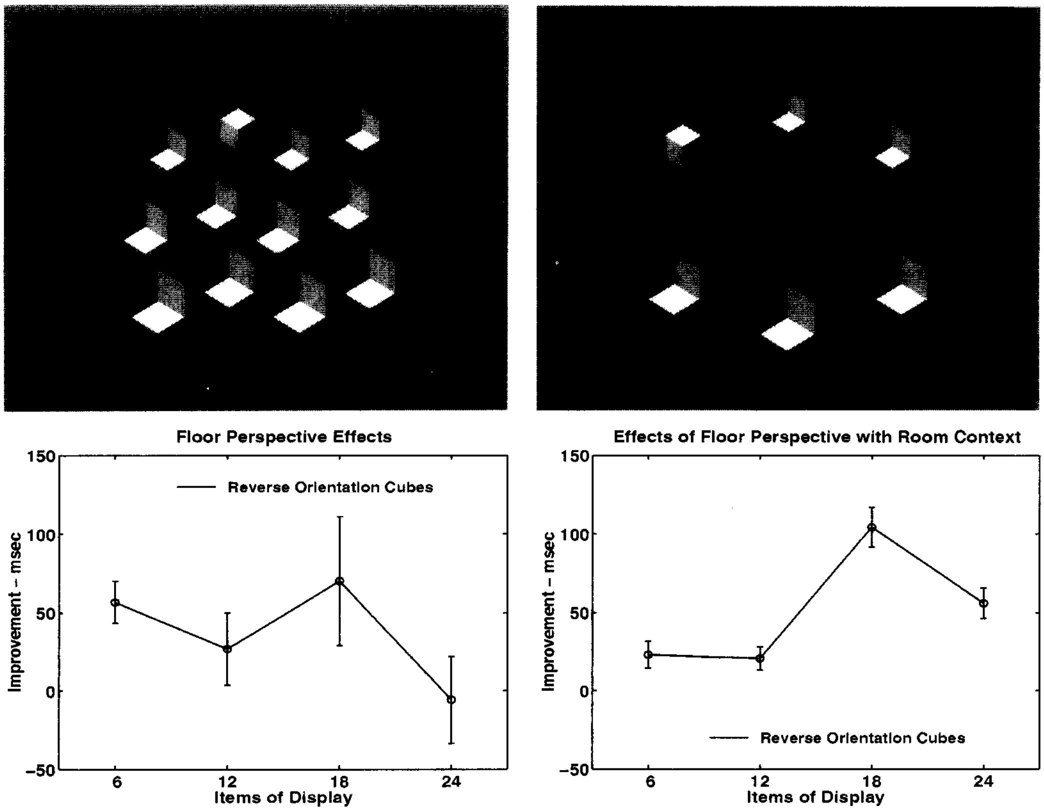
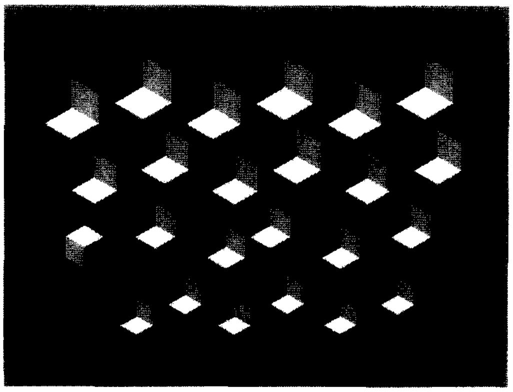


FIGURE 13. Reverse orientation cubes arranged in floor perspective (left) and with room context (right) are shown on the top row. Bottom graphs depict the respective effects.

cubes might be best perceived as bottom-lit cubes hanging from the ceiling, we tested three subjects on stimuli consisting of reversed cubes arranged in a ceiling perspective, as shown on the left of Fig. 14. Data were only collected for three display sizes: 12, 18 and 24 items. We see overall improvement that is correlated with subjects' reports of enhanced 3-D perception (Fig. 14, right). The confidence levels of improvement for the three increasing display sizes are >99%, >99%, and 97%.

Experiment 2(C). In this experiment, we extend our investigation of contextual effects to a pattern other than the shaded cube. For a rotated Y-junction in a square, which may be interpreted as a hole, we ask the question: how does a context that has either a consistent or an inconsistent 3-D interpretation with respect to its embedded patterns affect performance? Figure 15(A and B) shows the displays which have distractor holes that are respectively consistent and inconsistent with the 3-D wall frame. For control, an analogous surrounding frame that has no 3-D interpretation was used [Fig. 15(C)].

Data were collected from four subjects for the consistent 3-D frame experiment, and three subjects for the inconsistent 3-D frame experiment. Compared with the controls, the 3-D frame that was consistent with the distractor holes facilitated performance significantly, at a



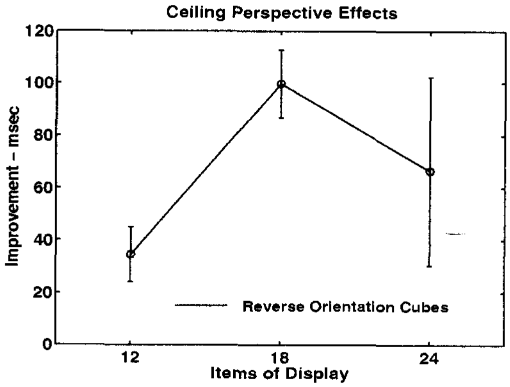


FIGURE 14. Reverse orientation cubes were also tested in ceiling perspective (left). The results are summarized on this graph in terms of improvement (right).

confidence level of >99% for every display size. There is also a trend for larger improvements to occur for larger display sizes (Fig. 16, left).

The 3-D frame inconsistent with the distractor holes did not lead to statistically significant improvements or impairments (Fig. 16, right). For the inconsistent frame case, some subjects saw the distractors as protruding cones, which would be consistent with the shading of the frame, instead of inconsistent holes. Other subjects saw the distractors as inconsistent holes only. We suspect that performance may have been facilitated for those who formed the consistent percept, but not for those who formed only the inconsistent percept. This dichotomy in perception might explain the large error bars.

Discussion

The results from all three sets of inducement experiments suggest that contextual information influences perception and performance. There are instances of improvement as well as impairment, with effects that are generally larger for the larger display sizes [Figs 10–12, Fig. 13 (left), Fig. 16 (right)]. We believe that these results can be best understood as a combination of bottom-up textural effects and top-down expectation effects.

Textural effects. Figure 10 shows a correlation between the disruption of textural uniformity and the breakdown of perceptual pop-out. This result suggests the involvement of textural mechanisms alongside the fast and parallel processing of 3-D shapes.

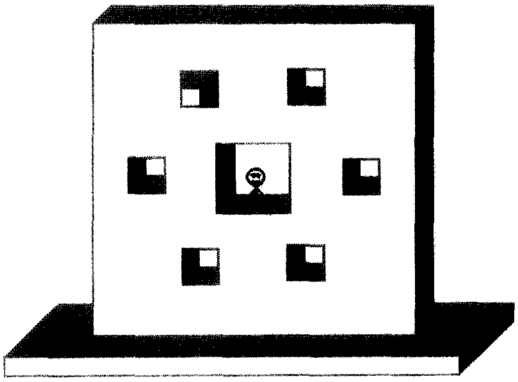
Dramatic impairment of performance is "rescued" to a large extent, however, if the different-sized cubes, instead of being positioned at random, are arranged according to size (Fig. 11, left). The largest increase in necessary SOA is about 30 msec instead of over 100 msec. We suggest that this partial rescue is due to the fact that the cubes, when arranged according to size, give rise to a texture that is, at least locally, homogeneous. Background homogeneity has been shown to have a large effect on search efficiency (Duncan & Humphreys, 1989; Wolfe *et al.*, 1992). The remaining impairment could be explained if the texture mechanism preferentially serves uniform textures, or if some of these cubes are of a size that hinders discrimination, either too big or too small.

Not only can the texture gradient perform partial rescue, we see from experiment 2(B) that it can even enhance performance in the difficult task involving reverse orientation cubes [Fig. 13 (left) and Fig. 14]. In particular, reverse cubes shown in a ceiling perspective resulted in a large improvement. For the 18-item display, the improvement is around 100 msec.

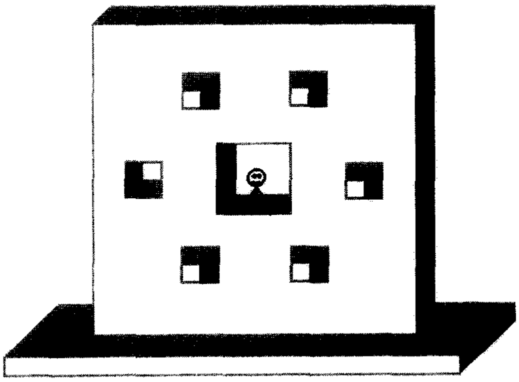
It is well known that texture density gradients can induce the percept of a receding plane (Gibson, 1950). We suggest that the textural mechanisms for the perception of surface slant are engaged here. When the reverse cubes are shown in perspective view, textural mechanisms for extracting ground-plane slant are driven by the apparent texture gradient. The resulting percept of a surface in 3-D would enhance the interpretation of the display as a physical scene, allowing the patterns to be perceived as 3-D shapes. Subjects' reports confirm that, indeed, the ceiling perspective enhanced the perception of the patterns as cubes hanging from the ceiling. For the floor perspective condition, some subjects described that the reverse orientation cubes looked like cubes balanced on a vertex.

In experiment 2(C), the effect of contextual information is extended to a stimulus other than the shaded cube. We find that a consistent wall frame improves performance while an inconsistent one does not. We suggest that textural effects may be at work here also. Textural density has been shown to be an important factor in texture segmentation experiments (Julesz, 1981). For the consistent-frame experiment, the upper-right corner of the wall has the exact same configuration as that of the target [Fig. 15(A)]. This corner junction adds to the textural density of the target pattern, and may increase its saliency, consequently improving performance.

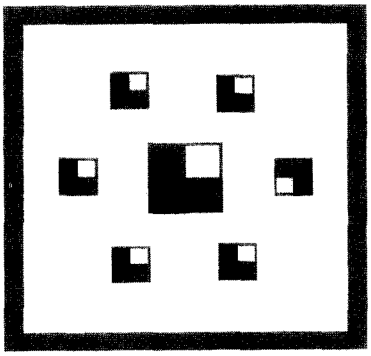
Expectation effects. Textural mechanisms, however, cannot account for all of our observed contextual effects. Evidence from experiments 2(A and B) points to an influence other than textural effects. Adding a background room context gave strong effects in all cases [Fig.



A



B



C

FIGURE 15. Consistent distractor holes in 3-D wall (A), inconsistent distractor holes in 3-D wall (B), and frame with no 3-D interpretation for control experiments (C).

11 (right), Fig. 12 and Fig. 13 (right)]. In Fig. 11, we see that while perspective cues alone counteracted somewhat the effect of differential sizing, the background room context actually led to significant improvement for some display sizes. Similarly, in experiment 2(B), the improvement went from being insignificant for all but the smallest display size to being statistically significant for every display size when the room background was added (Fig. 13). Improvements for the 18- and 24-item displays were particularly large.

Can these effects be explained also by bottom-up textural mechanisms? Perhaps the two shaded junctions that made up the back corners of the room provided additional textural density? This is unlikely because these junctions do not have the same configuration or shading as either the distractors or the target. Perhaps the luminance discontinuity formed by the top or the base of the two side walls, depending on the experiment, served as "guidelines" for perceiving the texture gradient, thereby facilitating the texture gradient mechanisms for extracting surface slant? Again, we believe this effect to be minimal at most, since these lines are quite short, spanning less than a quarter of the height of the display.

We suggest, as an alternative, top-down expectation effects. Both the background room context in experiments 2(A and B) and the wall frame in experiment 2(C) were statically displayed on the screen and did not flash on and off at short durations with the target and distractor stimuli. These static background displays may have served as a constant reminder that the stimuli about to be flashed on should be given a particular 3-D interpretation. When the flashed stimuli were consistent with the preconceived scene interpretation, perception was facilitated, and performance was enhanced [Fig. 11 (right), Fig. 13 (right) and Fig. 16 (left)]. On the other hand, when the flashed stimuli were inconsistent with the preconceived interpretation, performance was impaired (Fig. 12). This expectation effect may be related to the idea of top-down guidance in the "guided search model" proposed by Wolfe and collaborators. In the guided search model, attention can be guided in parallel by top-down information, allowing for increased search efficiency (Cave & Wolfe, 1990; Wolfe *et al.*, 1989).

GENERAL DISCUSSION

Our 2AFC short duration SOA experiments confirm Enns & Rensink's finding that three-dimensional shape from shading can be processed in parallel. We believe that this fast and parallel processing is dependent upon 3-D information because:

1. Shaded stimuli that are easily interpretable as familiar three dimensional shapes are processed fast and in parallel, while similar control stimuli that do not have such interpretation are not.
2. Distractor-target reversal experiments that are equivalent in two-dimensional space, differing only in their 3-D interpretations, show asymmetry in performance. This asymmetry is seen with the cubes as well as the Y-junction in circles.
3. 3-D contextual information can influence performance, both positively and negatively, depending on the degree to which the context contributes to a consistent 3-D interpretation, as suggested by the results of experiments 2(A, B and C).
4. Subjects' reports of 3-D perception coincide with performance that indicates fast, parallel processing.
5. Moreover, our results suggest that this parallel 3-D process has the following characteristics.

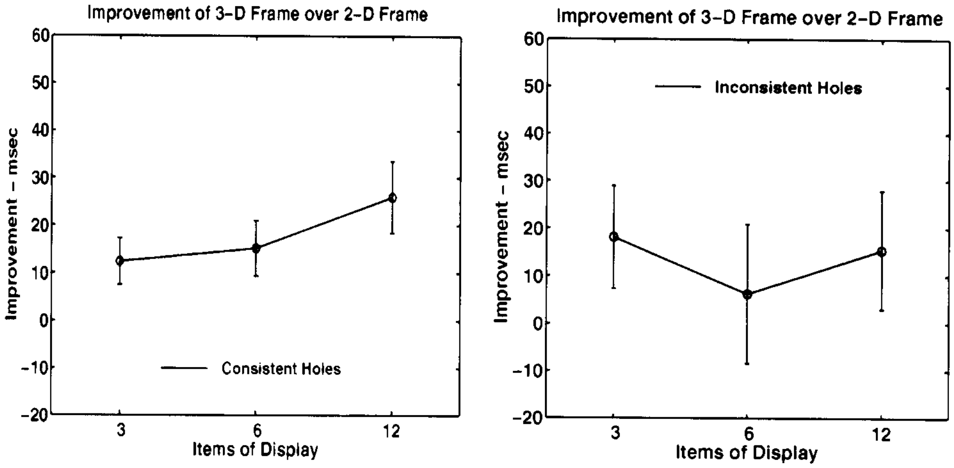


FIGURE 16. Improvement of performance for the consistent frame condition over the control condition (left) and improvement of performance for the inconsistent frame condition (right).

Has fast processing times

For the normal orientation shaded cubes and pies, our experiments yielded necessary SOA durations between 30 and 80 msec. These results suggest fast processing times for shaded 3-D stimuli, comparable to the ones previously reported in the classical "pop-out" and texture segregation experiments conducted using 2-D stimuli (Bergen & Julesz, 1983; Kröse, 1987; Gurnsey & Browse, 1987; Nothdurft, 1991).

Prefers shaded stimuli

Unlike the results reported by Enns & Rensink (1991), our results indicate that unshaded line stimuli do not drive this fast and parallel process. They are processed more slowly and more serially. Other experimental results also support our finding that shading is a crucial component for 3-D pop-out; shaded bubbles, which contain no internal line edges, are found to be processed in parallel also (Braun, 1990, 1993).

Computes locally on the Y-junction

The normal orientation shaded Y-junction is a salient cue recognized by this 3-D process. Results from experiments 1(C and E) suggest that computation begins locally at the Y-junction, and that perception of a complete 3-D solid is not necessary.

Subserves familiar shapes

Familiar shapes in generic views drive this process better than unfamiliar ones. This is evidenced by the asymmetry experiments in experiment 1(D), as well as experiment 1(C). Convex, top-lit shapes, are processed with ease, while concave or bottom-lit shapes are not. Generic views of familiar shapes, such as the cube, are preferred. A similar positive effect of familiarity on search tasks concerning 2-D line patterns was reported recently by Wang *et al.* (1994). They also found that performance is better when the distractors are familiar, than when the targets are.

Is influenced by contextual information

Our second set of experiments show that this 3-D process can be influenced by contextual information. Consistent contextual information that enhances the perception of a 3-D scene facilitates this process and improves performance, and inconsistent contextual information can impair performance. We suggest that these influences are mediated by bottom-up textural mechanisms as well as top-down expectation effects.

REFERENCES

Beck, J. (1966). Effects of orientation and of shape similarity on perceptual grouping. *Perception & Psychophysics, 1*, 300–302.

Beck, J. (1967). Perceptual grouping produced by line figures. *Perception & Psychophysics, 2*, 491–495.

Beck, J. (1982). Textural segmentation. In J. Beck (Ed.), *Organization and representation in perception* (pp. 285-317). Hillsdale, NJ: Lawrence Erlbaum.

Bergen, J. R. & Julesz, B. (1983). Parallel versus serial processing in rapid pattern discrimination. *Nature*, 696–698.

Braun, J. (1990). Focal attention and shape-from-shading. *Perception, 19*, A112.

Braun, J. (1993). Shape-from-shading is independent of visual attention and may be a "texton". *Spatial Vision, 74*, 311–322.

Cave, K. R. & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology, 22*, 225–271.

Duncan, J. & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 963*, 433–458.

Enns, J. T. & Rensink, R. A. (1990). Influence of scene-based properties on visual search. *Science, 247*, 721–723.

Enns, J. T. & Rensink, R. A. (1991). Preattentive recovery of three-dimensional orientation from line drawings. *Psychological Review, 983*, 335–351.

Gibson, J. (1950). *The perception of the visual world.* Boston, MA: Houghton Mifflin.

Gurnsey, R. & Browse, R. A. (1987). Micropattern properties and presentation conditions influencing visual texture discrimination. *Perception and Psychophysics, 403*, 239–252.

He, Z. J. & Nakayama, K. (1992). Surfaces versus features in visual search. *Nature, 3596392*, 231–233.

He, Z. J. & Nakayama, K. (1993). Common surface rather than common depth determines attention in 3-D search task. *Society for Neuroscience Abstracts, 191*, 773.

Julesz, B. (1975). Binocular depth perception without familiarity cues. *Science, 145*, 356–362.

Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. *Nature, 290*, 91–97.

Julesz, B. (1981). A brief outline of the texton theory of human vision. *Trends in Neuroscience, 72*, 41–45.

Julesz, B. (1984). Toward an axiomatic theory of preattentive vision. In Edelman, G. M., Gall, W. E. & Cowan, M. (Eds), *Dynamic aspects of neocortical function* (pp. 585–611). New York: Neurosciences Research Foundation.

Kleffner, D. A. & Ramachandran, V. S. (1992). On the perception of shape from shading. *Perception and Psychophysics, 521*, 18–36.

Kröse, B. J. A. (1987). Local structure analysers as determinants of preattentive pattern discrimination. *Biological Cybernetics, 55*, 289–298.

McNichol, D. (1972). *A primer of signal detection theory*. London, U.K.: George Allen & Unwin.

Nakayama, K. & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science, 2575075*, 1357–1363.

Nothdurft, H. C. (1991). Texture segmentation and pop-out from orientation contrast. *Vision Research, 316*, 1073–1078.

Olson, R. & Attneave, F. (1970). What variables produce similarity grouping? American Journal of Psychology, 83, 1–21.

Ramachandran, V. S. (1988). Perception of shape from shading. *Nature, 3316152*, 163–166.

Rubenstein, B. & Sagi, D. (1990). Spatial variability as a limiting factor in texture-discrimination tasks—implications for performance asymmetries. *Journal of the Optical Society of America A, 79*, 1632–1643.

Sun, J. Y., Perona, P. (1990). Preattentive perception of elementary three dimensional shapes. *Investigative Ophthalmology & Visual Science, 344*, 1083.

Sun, J. Y. & Perona, P. (1993). Preattentive perception of elementary three dimensional shapes. CIT/CNS 29-93 California Institute of Technology, August, 1993.

Sun, J. Y. & Perona, P. (1995). Influence of 3-D shape interpretation on preattentive popout and 2-D mechanisms. *Investigative Ophthalmology & Visual Science, 364*, 902.

Sun, J. Y. & Perona, P. (1996). Early computation of shape and reflectance in the visual system. *Nature*, in press.

Symons, L. A., Humphrey, G. K., Herbert, A. M. & Goodale, M. A. (1993). Region segregation in a visual form agnosic. *Investigative Ophthalmology & Visual Science, 344*, 1084.

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception Performance, 8*, 194–214.

Treisman, A. & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12*, 97–136.

Treisman, A. & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review, 95*, 15–48.

Treisman, A. & Patterson, R. (1984). Emergent features, attention, and object perception. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 12–31.

Wang, Q. Q., Cavanagh, P. & Green, M. (1994). Familiarity and popout in visual-search. *Perception and Psychophysics, 565*, 495–500.

Williams, D. & Julesz, B. (1992). Perceptual asymmetry in texture perception. *Proceedings of the National Academy of Science USA, 89*, 6531–6534.

Wolfe, J. M., Cave, K. R. & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model of visual search. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 419–433.

Wolfe, J. M., Freidman-Hill, S. R., Stewart, M. I. & O'Connell, K. M. (1992). The role of categorization in visual search for orientation. *Journal of Experimental Psychology: Human Perception and Performance, 181*, 34–49.

Acknowledgements—We are particularly grateful to Dr Jochen Braun and Dr Takeo Watanabe for many valuable suggestions on improving an earlier version of this manuscript. We would also like to thank Drs John Allman, Bela Julesz, Jitendra Malik, Douglas Shy, and Petra Stoerig for their helpful comments and suggestions, and subject LH for his generous donation of time and effort. We would also like to thank the NSF, the National Eye Institute for the ARVO/NEI Travel Fellowship Grant to ARVO 1993 where these results were first presented, and the NSF Center for Neuromorphic Systems Engineering at Caltech.